A Mathematical Investigation of Hallucination and Creativity in GPT Models

Year: 2023 | Citations: 120 | Authors: Minhyeok Lee

Abstract

In this paper, we present a comprehensive mathematical analysis of the hallucination phenomenon in generative pretrained transformer (GPT) models. We rigorously define and measure hallucination and creativity using concepts from probability theory and information theory. By introducing a parametric family of GPT models, we characterize the trade-off between hallucination and creativity and identify an optimal balance that maximizes model performance across various tasks. Our work offers a novel mathematical framework for understanding the origins and implications of hallucination in GPT models and paves the way for future research and development in the field of large language models (LLMs).